ABSTRACT

A virtualization infrastructure that allows multiple guest partitions to run within a host hardware partition. The host system is divided into distinct logical or virtual partitions and special infrastructure partitions are implemented to control resource management and to control physical I/O device drivers that are, in turn, used by operating systems in other distinct logical or virtual guest partitions. Host hardware resource management runs as a tracking application in a resource management "ultravisor" partition, while host resource management decisions are performed in a higher level command partition based on policies maintained in a separate operations partition. The conventional hypervisor is reduced to a context switching and containment element (monitor) for the respective partitions, while the system resource management functionality is implemented in the ultravisor partition. The ultravisor partition maintains the master in-memory database of the hardware resource allocations and serves a command channel to accept transactional requests for assignment of resources to partitions. It also provides individual read-only views of individual partitions to the associated partition monitors. Host hardware I/O management is implemented in special redundant I/O partitions. A scalable partition memory mapping system is implemented in the ultravisor partition so that the virtualized system is scalable to a virtually unlimited number of pages. A log (2^{10}) based allocation allows the virtual partition memory sizes to grow over multiple generations without increasing the overhead of managing the memory allocations. Each page of memory is assigned to one partition descriptor in the page hierarchy and is managed by the ultravisor partition.